

# State Complexity of Two Combined Operations: Reversal-Catenation and Star-Catenation

Bo Cui, Yuan Gao, Lila Kari, and Sheng Yu

June 25, 2010

## Abstract

In this paper, we show that, due to the structural properties of the resulting automaton obtained from a prior operation, the state complexity of a combined operation may not be equal but close to the mathematical composition of the state complexities of its component operations. In particular, we provide two witness combined operations: reversal combined with catenation and star combined with catenation.

## 1 Introduction

State complexity is a type of descriptional complexity based on *deterministic finite automaton* (DFA) model. The state complexity of an operation on regular languages is the number of states that are necessary and sufficient in the worst case for the minimal, complete DFA that accepts the resulting language of the operation. While many results on the state complexities of individual operations, such as union, intersection, catenation, star, reversal, shuffle, orthogonal catenation, proportional removal, and cyclic shift [2, 5, 6, 12, 14, 16, 15, 21, 23, 25], have been obtained in the past 15 years, the research of state complexities of combined operations, which was initiated by A. Salomaa, K. Salomaa, and S. Yu in 2007 [20], is attracting more attention. This is because, in practice, a combination of several individual operations, rather than only one individual operation, is often performed in a certain order. For example, in order to obtain a precise regular expression, a combination of basic operations is usually required.

In recent publications [3, 4, 8, 9, 10, 11, 17, 18, 20], it has been shown that the state complexity of a combined operation is not always a simple mathematical composition of the state complexities of its component operations. This is sometimes due to the structural properties of the DFA accepting the resulting language obtained from a prior operation of a combined operation. For example, the languages that are obtained from performing reversal and reach the upper bound of the state complexity of this operation are accepted by DFAs such that half of their states are final; and the initial state of the DFA accepting a language obtained after performing star is always a final state. As a result,

the resulting language obtained from a prior operation may not be among the worst cases of the subsequent operation. Since such issues are not concerned by the study of the state complexity of individual operations, they are certainly important in the research of the state complexity of combined operations. Although the number of combined operations is unlimited and it is impossible to study the state complexities of all of them, the study on combinations of two individual operations is clearly necessary.

In this paper, we study the state complexities of reversal combined with catenation, i.e.,  $L(A)^R L(B)$ , and star combined with catenation, i.e.,  $L(A)^* L(B)$ , for minimal complete DFAs  $A$  and  $B$  of sizes  $m, n \geq 1$ , respectively. For  $L(A)^R L(B)$ , we will show that the general upper bound  $\frac{3}{4}2^{m+n}$ , which is close to the composition of the state complexities of reversal and catenation  $2^{m+n} - 2^{n-1}$ , is reachable when  $m, n \geq 2$ , and it can be lower to  $2^{n-1}$  and  $2^{m-1} + 1$  when  $m = 1$  and  $n \geq 1$  and when  $m \geq 2$  and  $n = 1$ , respectively. For  $L(A)^* L(B)$ , we will show that, if  $A$  has only one final state and it is also the initial state, i.e.,  $L(A) = L(A)^*$ , the state complexity of catenation (also  $L(A)^* L(B)$ ) is  $m(2^n - 1) - 2^{n-1} + 1$ , which is lower than that of catenation  $m2^n - 2^{n-1}$ . In the other cases, that is when  $A$  contains some final states that are not the initial state, the state complexity of  $L(A)^* L(B)$  is  $5 \cdot 2^{m+n-3} - 2^{m-1} - 2^n + 1$  instead of  $\frac{3}{4}2^{m+n} - 2^{n-1}$ , the composition of the state complexities of star and catenation.

In the next section, we introduce the basic definitions and notations used in the paper. Then, we prove our results on reversal combined with catenation and star combined with catenation in Sections 3 and 4, respectively. We conclude the paper in Section 5.

## 2 Preliminaries

A DFA is denoted by a 5-tuple  $A = (Q, \Sigma, \delta, s, F)$ , where  $Q$  is the finite set of states,  $\Sigma$  is the finite input alphabet,  $\delta : Q \times \Sigma \rightarrow Q$  is the state transition function,  $s \in Q$  is the initial state, and  $F \subseteq Q$  is the set of final states. A DFA is said to be complete if  $\delta(q, a)$  is defined for all  $q \in Q$  and  $a \in \Sigma$ . All the DFAs we mention in this paper are assumed to be complete. We extend  $\delta$  to  $Q \times \Sigma^* \rightarrow Q$  in the usual way.

A *non-deterministic finite automaton* (NFA) is denoted by a 5-tuple  $A = (Q, \Sigma, \delta, s, F)$ , where the definitions of  $Q$ ,  $\Sigma$ ,  $s$ , and  $F$  are the same to those of DFAs, but the state transition function  $\delta$  is defined as  $\delta : Q \times \Sigma \rightarrow 2^Q$ , where  $2^Q$  denotes the power set of  $Q$ , i.e. the set of all subsets of  $Q$ .

In this paper, the state transition function  $\delta$  is often extended to  $\hat{\delta} : 2^Q \times \Sigma \rightarrow 2^Q$ . The function  $\hat{\delta}$  is defined by  $\hat{\delta}(R, a) = \{\delta(r, a) \mid r \in R\}$ , for  $R \subseteq Q$  and  $a \in \Sigma$ . We just write  $\delta$  instead of  $\hat{\delta}$  if there is no confusion.

A word  $w \in \Sigma^*$  is accepted by a finite automaton if  $\delta(s, w) \cap F \neq \emptyset$ . Two states in a finite automaton  $A$  are said to be *equivalent* if and only if for every word  $w \in \Sigma^*$ , if  $A$  is started in either state with  $w$  as input, it either accepts in both cases or rejects in both cases. It is well-known that a language which is accepted by an NFA can be accepted by a DFA, and such a language is said

to be *regular*. The language accepted by a DFA  $A$  is denoted by  $L(A)$ . The reader may refer to [13, 24] for more details about regular languages and finite automata.

The *state complexity* of a regular language  $L$ , denoted by  $sc(L)$ , is the number of states of the minimal complete DFA that accepts  $L$ . The state complexity of a class  $S$  of regular languages, denoted by  $sc(S)$ , is the supremum among all  $sc(L)$ ,  $L \in S$ . The state complexity of an operation on regular languages is the state complexity of the resulting languages from the operation as a function of the state complexity of the operand languages. Thus, in a certain sense, the state complexity of an operation is a worst-case complexity.

### 3 Reversal combined with catenation

In this section, we study the state complexity of  $L_1^R L_2$  for an  $m$ -state DFA language  $L_1$  and an  $n$ -state DFA language  $L_2$ . We first show that the state complexity of  $L_1^R L_2$  is upper bounded by  $\frac{3}{4}2^{m+n}$  in general (Theorem 1). Then we prove that this upper bound can be reached when  $m, n \geq 2$  (Theorem 2). Next, we investigate the case when  $m = 1$  and  $n \geq 1$  and prove the state complexity can be lower to  $2^{n-1}$  in such a case (Theorem 4). Finally, we show that the state complexity of  $L_1^R L_2$  is  $2^{m-1} + 1$  when  $m \geq 2$  and  $n = 1$  (Theorem 7).

Now, we start with a general upper bound of state complexity of  $L_1^R L_2$  for any integers  $m, n \geq 1$ .

**Theorem 1.** *For two integers  $m, n \geq 1$ , let  $L_1$  and  $L_2$  be two regular languages accepted by an  $m$ -state DFA and an  $n$ -state DFA, respectively. Then there exists a DFA of at most  $\frac{3}{4}2^{m+n}$  states that accepts  $L_1^R L_2$ .*

*Proof.* Let  $M = (Q_M, \Sigma, \delta_M, s_M, F_M)$  be a DFA of  $m$  states,  $k_1$  final states and  $L_1 = L(M)$ . Let  $N = (Q_N, \Sigma, \delta_N, s_N, F_N)$  be another DFA of  $n$  states and  $L_2 = L(N)$ .

Let  $M' = (Q_M, \Sigma, \delta_{M'}, F_M, \{s_M\})$  be an NFA with  $k_1$  initial states.  $\delta_{M'}(p, a) = q$  if  $\delta_M(q, a) = p$  where  $a \in \Sigma$  and  $p, q \in Q_M$ . Clearly,

$$L(M') = L(M)^R = L_1^R.$$

By performing subset construction on NFA  $M'$ , we can get an equivalent,  $2^m$ -state DFA  $A = (Q_A, \Sigma, \delta_A, s_A, F_A)$  such that  $L(A) = L_1^R$ . Since  $M'$  has only one final state  $s_M$ , we know that  $F_A = \{i \mid i \subseteq Q_M, s_M \in i\}$ . Thus,  $A$  has  $2^{m-1}$  final states in total. Now we construct a DFA  $B = (Q_B, \Sigma, \delta_B, s_B, F_B)$  accepting the language  $L_1^R L_2$ , where

$$\begin{aligned} Q_B &= \{\langle i, j \rangle \mid i \in Q_A, j \subseteq Q_N\}, \\ s_B &= \langle s_A, \emptyset \rangle, \text{ if } s_A \notin F_A; \\ &= \langle s_A, \{s_N\} \rangle, \text{ otherwise,} \\ F_B &= \{\langle i, j \rangle \in Q_B \mid j \cap F_N \neq \emptyset\}, \\ \delta_B(\langle i, j \rangle, a) &= \langle i', j' \rangle, \text{ if } \delta_A(i, a) = i', \delta_N(j, a) = j', a \in \Sigma, i' \notin F_A; \\ &= \langle i', j' \cup \{s_N\} \rangle, \text{ if } \delta_A(i, a) = i', \delta_N(j, a) = j', a \in \Sigma, i' \in F_A. \end{aligned}$$

From the above construction, we can see that all the states in  $B$  starting with  $i \in F_A$  must end with  $j$  such that  $s_N \in j$ . There are in total  $2^{m-1} \cdot 2^{n-1}$  states which don't meet this.

Thus, the number of states of the minimal DFA accepting  $L_1^R L_2$  is no more than

$$2^{m+n} - 2^{m-1} \cdot 2^{n-1} = \frac{3}{4}2^{m+n}.$$

□

This result gives an upper bound for the state complexity of  $L_1^R L_2$ . Next we show that this bound is reachable when  $m, n \geq 2$ .

**Theorem 2.** *Given two integers  $m, n \geq 2$ , there exists a DFA  $M$  of  $m$  states and a DFA  $N$  of  $n$  states such that any DFA accepting  $L(M)^R L(N)$  needs at least  $\frac{3}{4}2^{m+n}$  states.*

*Proof.* Let  $M = (Q_M, \Sigma, \delta_M, 0, \{m-1\})$  be a DFA, shown in Figure 1, where  $Q_M = \{0, 1, \dots, m-1\}$ ,  $\Sigma = \{a, b, c, d\}$ , and the transitions are given as:

- $\delta_M(i, a) = i + 1 \bmod m$ ,  $i = 0, \dots, m-1$ ,
- $\delta_M(i, b) = i$ ,  $i = 0, \dots, m-2$ ,  $\delta_M(m-1, b) = m-2$ ,
- $\delta_M(m-2, c) = m-1$ ,  $\delta_M(m-1, c) = m-2$ ,  
if  $m \geq 3$ ,  $\delta_M(i, c) = i$ ,  $i = 0, \dots, m-3$ ,
- $\delta_M(i, d) = i$ ,  $i = 0, \dots, m-1$ ,

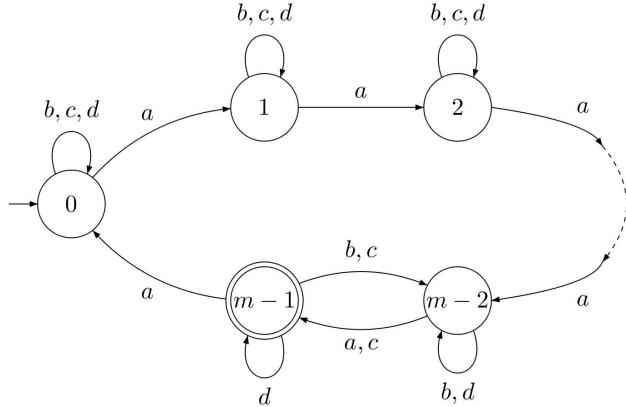


Figure 1: Witness DFA  $M$  of Theorem 2 showing that the upper bound in Theorem 1 is reachable when  $m, n \geq 2$

Let  $N = (Q_N, \Sigma, \delta_N, 0, \{n-1\})$  be a DFA, shown in Figure 2, where  $Q_N = \{0, 1, \dots, n-1\}$ ,  $\Sigma = \{a, b, c, d\}$ , and the transitions are given as:

- $\delta_N(i, a) = i, i = 1, \dots, n - 1,$
- $\delta_N(i, b) = i, i = 1, \dots, n - 1,$
- $\delta_N(i, c) = 0, i = 1, \dots, n - 1,$
- $\delta_N(i, d) = i + 1 \bmod n, i = 0, \dots, n - 1,$

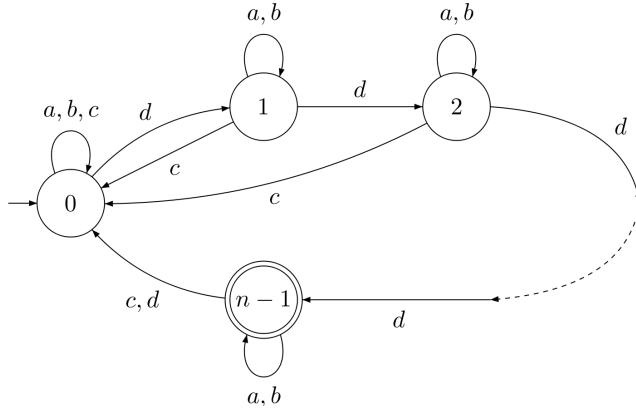


Figure 2: Witness DFA  $N$  of Theorem 2 showing that the upper bound in Theorem 1 is reachable when  $m, n \geq 2$

Now we design a DFA  $A = (Q_A, \Sigma, \delta_A, \{m-1\}, F_A)$ , where  $Q_A = \{q \mid q \subseteq Q_M\}$ ,  $\Sigma = \{a, b, c, d\}$ ,  $F_A = \{q \mid 0 \in q, q \in Q_A\}$ , and the transitions are defined as:

$$\delta_A(p, e) = \{j \mid \delta_M(j, e) = i, i \in p\}, p \in Q_A, e \in \Sigma.$$

It is easy to see that  $A$  is a DFA that accepts  $L(M)^R$ . We prove that  $A$  is minimal before using it.

(I) We first show that every state  $I \in Q_A$  is reachable from  $\{m-1\}$ . There are three cases.

1.  $|I| = 0$ .  $|I| = 0$  if and only if  $I = \emptyset$ .  $\delta_A(\{m-1\}, b) = I = \emptyset$ .
2.  $|I| = 1$ . Let  $I = \{i\}, 0 \leq i \leq m-1$ .  $\delta_A(\{m-1\}, a^{m-1-i}) = I$ .
3.  $2 \leq |I| \leq m$ . Let  $I = \{i_1, i_2, \dots, i_k\}, 0 \leq i_1 < i_2 < \dots < i_k \leq m-1, 2 \leq k \leq m$ .  $\delta_A(\{m-1\}, w) = I$ , where

$$w = ab(ac)^{i_2-i_1-1}ab(ac)^{i_3-i_2-1}\dots ab(ac)^{i_k-i_{k-1}-1}a^{m-1-i_k}.$$

(II) Any two different states  $I$  and  $J$  in  $Q_A$  are distinguishable.

Without loss of generality, we may assume that  $|I| \geq |J|$ . Let  $x \in I - J$ . Then a string  $a^x$  can distinguish these two states because

$$\begin{aligned}\delta_A(I, a^x) &\in F_A, \\ \delta_A(J, a^x) &\notin F_A.\end{aligned}$$

Due to (I) and (II),  $A$  is a minimal DFA with  $2^m$  states which accepts  $L(M)^R$ . Now let  $B = (Q_B, \Sigma, \delta_B, s_B, F_B)$  be another DFA, where

$$\begin{aligned}Q_B &= \{\langle p, q \rangle \mid p \in Q_A - F_A, q \subseteq Q_N\} \\ &\quad \cup \{\langle p', q' \rangle \mid p' \in F_A, q' \subseteq Q_N, 0 \in q'\}, \\ \Sigma &= \{a, b, c, d\}, \\ s_B &= \langle \{m-1\}, \emptyset \rangle, \\ F_B &= \{\langle p, q \rangle \mid n-1 \in q, \langle p, q \rangle \in Q_B\},\end{aligned}$$

and for each state  $\langle p, q \rangle \in Q_B$  and each letter  $e \in \Sigma$ ,

$$\delta_B(\langle p, q \rangle, e) = \begin{cases} \langle p', q' \rangle & \text{if } \delta_A(p, e) = p' \notin F_A, \delta_N(q, e) = q', \\ \langle p', q' \rangle & \text{if } \delta_A(p, e) = p' \in F_A, \delta_N(q, e) = r', q' = r' \cup \{0\}. \end{cases}$$

As we mentioned in last proof, all the states starting with  $p \in F_A$  must end with  $q \subseteq Q_N$  such that  $0 \in q$ . Clearly,  $B$  accepts the language  $L(M)^R L(N)$  and it has

$$2^m \cdot 2^n - 2^{m-1} \cdot 2^{n-1} = \frac{3}{4}2^{m+n}$$

states. Now we show that  $B$  is a minimal DFA.

(I) Every state  $\langle p, q \rangle \in Q_B$  is reachable. We consider the following five cases:

1.  $p = \emptyset, q = \emptyset$ .  $\langle \emptyset, \emptyset \rangle$  is the sink state of  $B$ .  $\delta_B(\langle \{m-1\}, \emptyset \rangle, b) = \langle p, q \rangle$ .
2.  $p \neq \emptyset, q = \emptyset$ . Let  $p = \{p_1, p_2, \dots, p_k\}$ ,  $1 \leq p_1 < p_2 < \dots < p_k \leq m-1$ ,  $1 \leq k \leq m-1$ . Note that  $0 \notin p$ , because  $0 \in p$  guarantees  $0 \in q$ .  $\delta_B(\langle \{m-1\}, \emptyset \rangle, w) = \langle p, q \rangle$ , where

$$w = ab(ac)^{p_2-p_1-1}ab(ac)^{p_3-p_2-1} \dots ab(ac)^{p_k-p_{k-1}-1}a^{m-1-p_k}.$$

Please note that  $w = a^{m-1-p_1}$  when  $k = 1$ .

3.  $p = \emptyset, q \neq \emptyset$ . In this case, let  $q = \{q_1, q_2, \dots, q_l\}$ ,  $0 \leq q_1 < q_2 < \dots < q_l \leq n-1$ ,  $1 \leq l \leq n$ .  $\delta_B(\langle \{m-1\}, \emptyset \rangle, x) = \langle p, q \rangle$ , where

$$x = a^m d^{q_l-q_{l-1}} a^m d^{q_{l-1}-q_{l-2}} \dots a^m d^{q_2-q_1} a^m d^{q_1} b.$$

4.  $p \neq \emptyset, 0 \notin p, q \neq \emptyset$ . Let  $p = \{p_1, p_2, \dots, p_k\}$ ,  $1 \leq p_1 < p_2 < \dots < p_k \leq m-1$ ,  $1 \leq k \leq m-1$  and  $q = \{q_1, q_2, \dots, q_l\}$ ,  $0 \leq q_1 < q_2 < \dots < q_l \leq n-1$ ,  $1 \leq l \leq n$ . We can find a string  $uv$  such that  $\delta_B(\langle \{m-1\}, \emptyset \rangle, uv) = \langle p, q \rangle$ , where

$$u = ab(ac)^{p_2-p_1-1}ab(ac)^{p_3-p_2-1} \dots ab(ac)^{p_k-p_{k-1}-1}a^{m-1-p_k},$$

$$v = a^m d^{q_l-q_{l-1}} a^m d^{q_{l-1}-q_{l-2}} \dots a^m d^{q_2-q_1} a^m d^{q_1}.$$

5.  $p \neq \emptyset$ ,  $0 \in p$ ,  $m-1 \notin p$ ,  $q \neq \emptyset$ . Let  $p = \{p_1, p_2, \dots, p_k\}$ ,  $0 = p_1 < p_2 < \dots < p_k < m-1$ ,  $1 \leq k \leq m-1$  and  $q = \{q_1, q_2, \dots, q_l\}$ ,  $0 = q_1 < q_2 < \dots < q_l \leq n-1$ ,  $1 \leq l \leq n$ . Since  $0$  is in  $p$ , according to the definition of  $B$ ,  $0$  has to be in  $q$  as well. There exists a string  $u'v'$  such that  $\delta_B(\langle\{m-1\}, \emptyset\rangle, u'v') = \langle p, q \rangle$ , where

$$u' = ab(ac)^{p_2-p_1-1}ab(ac)^{p_3-p_2-1}\dots ab(ac)^{p_k-p_{k-1}-1}a^{m-2-p_k},$$

$$v' = a^m d^{q_l-q_{l-1}} a^m d^{q_{l-1}-q_{l-2}} \dots a^m d^{q_2-q_1} a^m d^{q_1} a.$$

6.  $p \neq \emptyset$ ,  $\{0, m-1\} \subseteq p$ ,  $q \neq \emptyset$ . Let  $p = \{p_1, p_2, \dots, p_k\}$ ,  $0 = p_1 < p_2 < \dots < p_k = m-1$ ,  $2 \leq k \leq m$  and  $q = \{q_1, q_2, \dots, q_l\}$ ,  $0 = q_1 < q_2 < \dots < q_l \leq n-1$ ,  $1 \leq l \leq n$ . In this case, we have

$$\langle p, q \rangle = \begin{cases} \delta_B(\langle\{0, 1, p_2+1, \dots, p_{k-1}+1\}, q\rangle, a), & \text{if } m-2 \notin p, \\ \delta_B(\langle p - \{m-1\}, q \rangle, b), & \text{if } m-2 \in p, \end{cases}$$

where states  $\langle\{0, 1, p_2+1, \dots, p_{k-1}+1\}, q\rangle$  and  $\langle p - \{m-1\}, q \rangle$  have been proved to be reachable in Case 5.

(II) We then show that any two different states  $\langle p_1, q_1 \rangle$  and  $\langle p_2, q_2 \rangle$  in  $Q_B$  are distinguishable.

1.  $q_1 \neq q_2$ . Without loss of generality, we may assume that  $|q_1| \geq |q_2|$ . Let  $x \in q_1 - q_2$ . A string  $d^{n-1-x}$  can distinguish them because

$$\begin{aligned} \delta_B(\langle p_1, q_1 \rangle, d^{n-1-x}) &\in F_B, \\ \delta_B(\langle p_2, q_2 \rangle, d^{n-1-x}) &\notin F_B. \end{aligned}$$

2.  $p_1 \neq p_2$ ,  $q_1 = q_2$ . Without loss of generality, we assume that  $|p_1| \geq |p_2|$ . Let  $y \in p_1 - p_2$ . Then there always exists a string  $a^y c^2 d^n$  such that

$$\begin{aligned} \delta_B(\langle p_1, q_1 \rangle, a^y c^2 d^n) &\in F_B, \\ \delta_B(\langle p_2, q_2 \rangle, a^y c^2 d^n) &\notin F_B. \end{aligned}$$

Since all the states in  $B$  are reachable and pairwise distinguishable, DFA  $B$  is minimal. Thus, any DFA accepting  $L(M))^R L(N)$  needs at least  $\frac{3}{4}2^{m+n}$  states.  $\square$

This result gives a lower bound for the state complexity of  $L_1^R L_2$  when  $m, n \geq 2$ . It coincides with the upper bound shown in Theorem 1 exactly. Thus, we obtain the state complexity of the combined operation  $L_1^R L_2$  for  $m \geq 2$  and  $n \geq 2$ .

**Theorem 3.** *For any integers  $m, n \geq 2$ , let  $L_1$  be an  $m$ -state DFA language and  $L_2$  be an  $n$ -state DFA language. Then  $\frac{3}{4}2^{m+n}$  states are both necessary and sufficient in the worst case for a DFA to accept  $L_1^R L_2$ .*

In the rest of this section, we study the remaining cases when either  $m = 1$  or  $n = 1$ .

We first consider the case when  $m = 1$  and  $n \geq 2$ . In this case,  $L_1 = \emptyset$  or  $L_1 = \Sigma^*$ .  $L_1^R L_2 = L_1 L_2$  holds no matter  $L_1$  is  $\emptyset$  or  $\Sigma^*$ , since  $\emptyset^R = \emptyset$  and  $(\Sigma^*)^R = \Sigma^*$ . It has been shown in [23] that  $2^{n-1}$  states are both sufficient and necessary in the worst case for a DFA to accept the catenation of a 1-state DFA language and an  $n$ -state DFA language,  $n \geq 2$ .

When  $m = 1$  and  $n = 1$ , it is also easy to see that 1 state is sufficient and necessary in the worst case for a DFA to accept  $L_1^R L_2$ , because  $L_1^R L_2$  is either  $\emptyset$  or  $\Sigma^*$ . Thus, we have the following theorem concerning the state complexity of  $L_1^R L_2$  for  $m = 1$  and  $n \geq 1$ .

**Theorem 4.** *Let  $L_1$  be a 1-state DFA language and  $L_2$  be an  $n$ -state DFA language,  $n \geq 1$ . Then  $2^{n-1}$  states are both sufficient and necessary in the worst case for a DFA to accept  $L_1^R L_2$ .*

Now, we study the state complexity of  $L_1^R L_2$  for  $m \geq 2$  and  $n = 1$ . Let us start with the following upper bound.

**Theorem 5.** *For any integer  $m \geq 2$ , let  $L_1$  and  $L_2$  be two regular languages accepted by an  $m$ -state DFA and a 1-state DFA, respectively. Then there exists a DFA of at most  $2^{m-1} + 1$  states that accepts  $L_1^R L_2$ .*

*Proof.* Let  $M = (Q_M, \Sigma, \delta_M, s_M, F_M)$  be a DFA of  $m$  states,  $m \geq 2$ ,  $k_1$  final states and  $L_1 = L(M)$ . Let  $N$  be another DFA of 1 state and  $L_2 = L(N)$ . Since  $N$  is a complete DFA, as we mentioned before,  $L(N)$  is either  $\emptyset$  or  $\Sigma^*$ . Clearly,  $L_1^R \cdot \emptyset = \emptyset$ . Thus, we need to consider only the case  $L_2 = L(N) = \Sigma^*$ .

We construct an NFA  $M' = (Q_M, \Sigma, \delta_{M'}, F_M, \{s_M\})$  with  $k_1$  initial states which is similar to the proof of Theorem 1.  $\delta_{M'}(p, a) = q$  if  $\delta_M(q, a) = p$  where  $a \in \Sigma$  and  $p, q \in Q_M$ . It is easy to see that

$$L(M') = L(M)^R = L_1^R.$$

By performing subset construction on NFA  $M'$ , we get an equivalent,  $2^m$ -state DFA  $A = (Q_A, \Sigma, \delta_A, s_A, F_A)$  such that  $L(A) = L_1^R$ .  $F_A = \{i \mid i \subseteq Q_M, s_M \in i\}$  because  $M'$  has only one final state  $s_M$ . Thus,  $A$  has  $2^{m-1}$  final states in total.

Define  $B = (Q_B, \Sigma, \delta_B, s_B, \{f_B\})$  where  $f_B \notin Q_A$ ,  $Q_B = (Q_A - F_A) \cup \{f_B\}$ ,

$$s_B = \begin{cases} s_A & \text{if } s_A \notin F_A, \\ f_B & \text{otherwise.} \end{cases}$$

and for any  $a \in \Sigma$  and  $p \in Q_B$ ,

$$\delta_B(p, a) = \begin{cases} \delta_A(p, a) & \text{if } \delta_A(p, a) \notin F_A, \\ f_B & \text{if } \delta_A(p, a) \in F_A, \\ f_B & \text{if } p = f_B. \end{cases}$$

The automaton  $B$  is exactly the same as  $A$  except that  $A$ 's  $2^{m-1}$  final states are made to be sink states and these sink, final states are merged into one,

since they are equivalent. When the computation reaches the final state  $f_B$ , it remains there. Now, it is clear that  $B$  has

$$2^m - 2^{m-1} + 1 = 2^{m-1} + 1$$

states and  $L(B) = L_1^R \Sigma^*$ . □

This theorem shows an upper bound for the state complexity of  $L_1^R L_2$  for  $m \geq 2$  and  $n = 1$ . Next we prove that this upper bound is reachable.

**Lemma 1.** *Given an integer  $m = 2$  or  $3$ , there exists an  $m$ -state DFA  $M$  and a 1-state DFA  $N$  such that any DFA accepting  $L(M)^R L(N)$  needs at least  $2^{m-1} + 1$  states.*

*Proof.* When  $m = 2$  and  $n = 1$ . We can construct the following witness DFAs. Let  $M = (\{0, 1\}, \Sigma, \delta_M, 0, \{1\})$  be a DFA, where  $\Sigma = \{a, b\}$ , and the transitions are given as:

- $\delta_M(0, a) = 1, \delta_M(1, a) = 0,$
- $\delta_M(0, b) = 0, \delta_M(1, b) = 0.$

Let  $N$  be the DFA accepting  $\Sigma^*$ . Then the resulting DFA for  $L(M)^R \Sigma^*$  is  $A = (\{0, 1, 2\}, \Sigma, \delta_A, 0, \{1\})$  where

- $\delta_A(0, a) = 1, \delta_A(1, a) = 1, \delta_A(2, a) = 2,$
- $\delta_A(0, b) = 2, \delta_A(1, b) = 1, \delta_A(2, b) = 2.$

When  $m = 3$  and  $n = 1$ . The witness DFAs are as follows. Let  $M' = (\{0, 1, 2\}, \Sigma', \delta_{M'}, 0, \{2\})$  be a DFA, where  $\Sigma' = \{a, b, c\}$ , and the transitions are:

- $\delta_{M'}(0, a) = 1, \delta_{M'}(1, a) = 2, \delta_{M'}(2, a) = 0,$
- $\delta_{M'}(0, b) = 0, \delta_{M'}(1, b) = 0, \delta_{M'}(2, b) = 1,$
- $\delta_{M'}(0, c) = 0, \delta_{M'}(1, c) = 2, \delta_{M'}(2, c) = 1.$

Let  $N'$  be the DFA accepting  $\Sigma'^*$ . The resulting DFA for  $L(M')^R \Sigma'^*$  is  $A' = (\{0, 1, 2, 3, 4\}, \Sigma', \delta_{A'}, 0, \{3\})$  where

- $\delta_{A'}(0, a) = 1, \delta_{A'}(1, a) = 3, \delta_{A'}(2, a) = 2, \delta_{A'}(3, a) = 3, \delta_{A'}(4, a) = 3,$
- $\delta_{A'}(0, b) = 2, \delta_{A'}(1, b) = 4, \delta_{A'}(2, b) = 2, \delta_{A'}(3, b) = 3, \delta_{A'}(4, b) = 4,$
- $\delta_{A'}(0, c) = 1, \delta_{A'}(1, c) = 0, \delta_{A'}(2, c) = 2, \delta_{A'}(3, c) = 3, \delta_{A'}(4, c) = 4.$

□

The above result shows that the bound  $2^{m-1} + 1$  is reachable when  $m$  is equal to 2 or 3 and  $n = 1$ . The last case is  $m \geq 4$  and  $n = 1$ .

**Theorem 6.** *Given an integer  $m \geq 4$ , there exists a DFA  $M$  of  $m$  states and a DFA  $N$  of 1 state such that any DFA accepting  $L(M)^R L(N)$  needs at least  $2^{m-1} + 1$  states.*

*Proof.* Let  $M = (Q_M, \Sigma, \delta_M, 0, \{m-1\})$  be a DFA, shown in Figure 3, where  $Q_M = \{0, 1, \dots, m-1\}$ ,  $m \geq 4$ ,  $\Sigma = \{a, b, c, d\}$ , and the transitions are given as:

- $\delta_M(i, a) = i + 1 \bmod m$ ,  $i = 0, \dots, m-1$ ,
- $\delta_M(i, b) = i$ ,  $i = 0, \dots, m-2$ ,  $\delta_M(m-1, b) = m-2$ ,
- $\delta_M(i, c) = i$ ,  $i = 0, \dots, m-3$ ,  $\delta_M(m-2, c) = m-1$ ,  $\delta_M(m-1, c) = m-2$ ,
- $\delta_M(0, d) = 0$ ,  $\delta_M(i, d) = i + 1$ ,  $i = 1, \dots, m-2$ ,  $\delta_M(m-1, d) = 1$ .

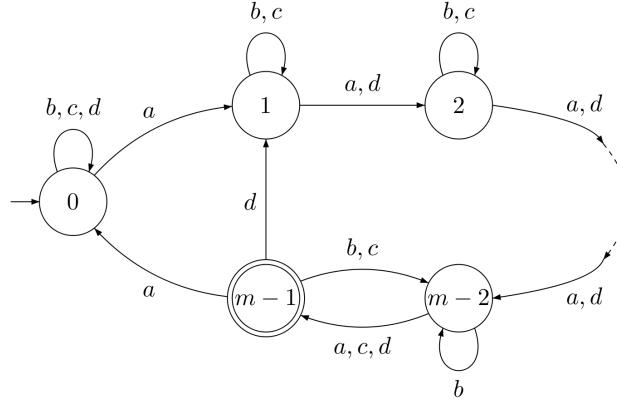


Figure 3: Witness DFA  $M$  of Theorem 6 showing that the upper bound in Theorem 5 is reachable when  $m \geq 4$  and  $n = 1$

Let  $N$  be the DFA accepting  $\Sigma^*$ . Then  $L(M)^R L(N) = L(M)^R \Sigma^*$ . Now we design a DFA  $A = (Q_A, \Sigma, \delta_A, \{m-1\}, F_A)$  similar to the proof of Theorem 2, where  $Q_A = \{q \mid q \subseteq Q_M\}$ ,  $\Sigma = \{a, b, c, d\}$ ,  $F_A = \{q \mid 0 \in q, q \in Q_A\}$ , and the transitions are defined as:

$$\delta_A(p, e) = \{j \mid \delta_M(j, e) = i, i \in p\}, p \in Q_A, e \in \Sigma.$$

It is easy to see that  $A$  is a DFA that accepts  $L(M)^R$ . Since the transitions of  $M$  on letters  $a$ ,  $b$ , and  $c$  are exactly the same as those of DFA  $M$  in the proof of Theorem 2, we can say that  $A$  is minimal and it has  $2^m$  states, among which  $2^{m-1}$  states are final.

Define  $B = (Q_B, \Sigma, \delta_B, s_B, \{f_B\})$  where  $f_B \notin Q_A$ ,  $Q_B = (Q_A - F_A) \cup \{f_B\}$ ,

$$s_B = \begin{cases} s_A & \text{if } s_A \notin F_A, \\ f_B & \text{otherwise.} \end{cases}$$

and for any  $e \in \Sigma$  and  $I \in Q_B$ ,

$$\delta_B(I, e) = \begin{cases} \delta_A(I, e) & \text{if } \delta_A(I, e) \notin F_A, \\ f_B & \text{if } \delta_A(I, e) \in F_A, \\ f_B & \text{if } I = f_B. \end{cases}$$

DFA  $B$  is the same as  $A$  except that  $A$ 's  $2^{m-1}$  final states are changed into sink states and merged to one sink, final state, as we did in the proof of Theorem 5. Clearly,  $B$  has  $2^m - 2^{m-1} + 1 = 2^{m-1} + 1$  states and  $L(B) = L(M)^R \Sigma^*$ . Next we show that  $B$  is a minimal DFA.

(I) Every state  $I \in Q_B$  is reachable from  $\{m-1\}$ . The proof is similar to that of Theorem 2. We consider the following four cases:

1.  $I = \emptyset$ .  $\delta_A(\{m-1\}, b) = I = \emptyset$ .
2.  $I = f_B$ .  $\delta_A(\{m-1\}, a^{m-1}) = I = f_B$ .
3.  $|I| = 1$ . Assume that  $I = \{i\}$ ,  $1 \leq i \leq m-1$ . Note that  $i \neq 0$  because all the final states in  $A$  have been merged into  $f_B$ . In this case,  $\delta_A(\{m-1\}, a^{m-1-i}) = I$ .
4.  $2 \leq |I| \leq m$ . Assume that  $I = \{i_1, i_2, \dots, i_k\}$ ,  $1 \leq i_1 < i_2 < \dots < i_k \leq m-1$ ,  $2 \leq k \leq m$ .  $\delta_A(\{m-1\}, w) = I$ , where

$$w = ab(ac)^{i_2-i_1-1}ab(ac)^{i_3-i_2-1}\dots ab(ac)^{i_k-i_{k-1}-1}a^{m-1-i_k}.$$

(II) Any two different states  $I$  and  $J$  in  $Q_B$  are distinguishable.

Since  $f_B$  is the only final state in  $Q_B$ , it is inequivalent to any other state. Thus, we consider the case when neither of  $I$  and  $J$  is  $f_B$ .

Without loss of generality, we may assume that  $|I| \geq |J|$ . Let  $x \in I - J$ .  $x$  is always greater than 0 because all the states which include 0 have been merged into  $f_B$ . Then a string  $d^{x-1}a$  can distinguish these two states because

$$\begin{aligned} \delta_B(I, d^{x-1}a) &= f_B, \\ \delta_B(J, d^{x-1}a) &\neq f_B. \end{aligned}$$

Since all the states in  $B$  are reachable and pairwise distinguishable,  $B$  is a minimal DFA. Thus, any DFA accepting  $L(M)^R \Sigma^*$  needs at least  $2^{m-1} + 1$  states.  $\square$

After summarizing Theorem 5, Theorem 6 and Lemma 1, we obtain the state complexity of the combined operation  $L_1^R L_2$  for  $m \geq 2$  and  $n = 1$ .

**Theorem 7.** *For any integer  $m \geq 2$ , let  $L_1$  be an  $m$ -state DFA language and  $L_2$  be a 1-state DFA language. Then  $2^{m-1} + 1$  states are both sufficient and necessary in the worst case for a DFA to accept  $L_1^R L_2$ .*

## 4 Star combined with catenation

In this section, we investigate the state complexity of  $L(A)^*L(B)$  for two DFAs  $A$  and  $B$  of sizes  $m, n \geq 1$ , respectively. We first notice that, when  $n = 1$ , the state complexity of  $L(A)^*L(B)$  is 1 for any  $m \geq 1$ . This is because  $B$  is complete ( $L(B)$  is either  $\emptyset$  or  $\Sigma^*$ ), and we have either  $L(A)^*L(B) = \emptyset$  or  $\Sigma^* \subseteq L(A)^*L(B) \subseteq \Sigma^*$ . Thus,  $L(A)^*L(B)$  is always accepted by a 1 state DFA. Next, we consider the case where  $A$  has only one final state and it is also the initial state. In such a case,  $L(A)^*$  is also accepted by  $A$ , and hence the state complexity of  $L(A)^*L(B)$  is equal to that of  $L(A)L(B)$ . We will show that, for any  $A$  of size  $m \geq 1$  in this form and any  $B$  of size  $n \geq 2$ , the state complexity of  $L(A)L(B)$  (also  $L(A)^*L(B)$ ) is  $m(2^n - 1) - 2^{n-1} + 1$  (Theorems 8 and 9), which is lower than the state complexity of catenation in the general case. Lastly, we consider the state complexity of  $L(A)^*L(B)$  in the remaining case, that is when  $A$  has at least a final state that is not the initial state and  $n \geq 2$ . We will show that its upper bound (Theorem 10) coincides with its lower bound (Theorem 11), and the state complexity is  $5 \cdot 2^{m+n-3} - 2^{m-1} - 2^n + 1$ .

Now, we consider the case where DFA  $A$  has only one final state and it is also the initial state, and first obtain the following upper bound of the state complexity of  $L(A)L(B)$  ( $L(A)^*L(B)$ ), for any DFA  $B$  of size  $n \geq 2$ .

**Theorem 8.** *For integers  $m \geq 1$  and  $n \geq 2$ , let  $A$  and  $B$  be two DFAs with  $m$  and  $n$  states, respectively, where  $A$  has only one final state and it is also the initial state. Then, there exists a DFA of at most  $m(2^n - 1) - 2^{n-1} + 1$  states that accepts  $L(A)L(B)$ , which is equal to  $L(A)^*L(B)$ .*

*Proof.* Let  $A = (Q_1, \Sigma, \delta_1, s_1, \{s_1\})$  and  $B = (Q_2, \Sigma, \delta_2, s_2, F_2)$ . We construct a DFA  $C = (Q, \Sigma, \delta, s, F)$  such that

$$\begin{aligned} Q &= Q_1 \times (2^{Q_2} - \{\emptyset\}) - \{s_1\} \times (2^{Q_2 - \{s_2\}} - \{\emptyset\}), \\ s &= \langle s_1, \{s_2\} \rangle, \\ F &= \{\langle q, T \rangle \in Q \mid T \cap F_2 \neq \emptyset\}, \\ \delta(\langle q, T \rangle, a) &= \langle q', T' \rangle, \text{ for } a \in \Sigma, \text{ where } q' = \delta_1(q, a) \text{ and } T' = R \cup \{s_2\} \\ &\quad \text{if } q' = s_1, T' = R \text{ otherwise, where } R = \delta_2(T, a). \end{aligned}$$

Intuitively,  $Q$  contains the pairs whose first component is a state of  $Q_1$  and second component is a subset of  $Q_2$ . Since  $s_1$  is the final state of  $A$ , without reading any letter, we can enter the initial state of  $B$ . Thus, states  $\langle q, \emptyset \rangle$  such that  $q \in Q_1$  can never be reached in  $C$ , because  $B$  is complete. Moreover,  $Q$  does not contain those states whose first component is  $s_1$  and second component does not contain  $s_2$ .

Clearly,  $C$  has  $m(2^n - 1) - 2^{n-1} + 1$  states, and we can verify that  $L(C) = L(A)L(B)$ .  $\square$

Next, we show that this upper bound can be reached by some witness DFAs in the specific form.

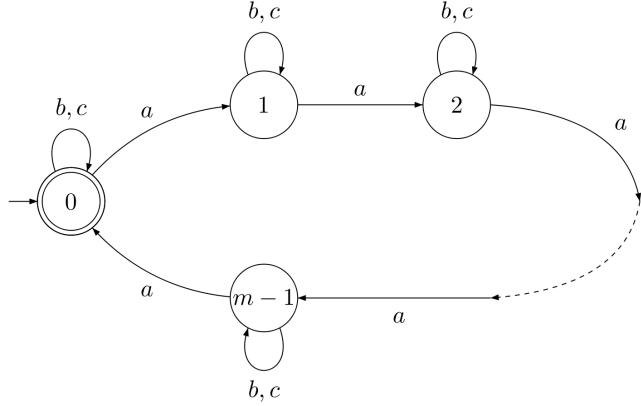


Figure 4: Witness DFA  $A$  for Theorem 9 when  $m \geq 2$

**Theorem 9.** *For any integers  $m \geq 1$  and  $n \geq 2$ , there exist a DFA  $A$  of  $m$  states and a DFA  $B$  of  $n$  states, where  $A$  has only one final state and it is also the initial state, such that any DFA accepting the language  $L(A)L(B)$ , which is equal to  $L(A)^*L(B)$ , needs at least  $m(2^n - 1) - 2^{n-1} + 1$  states.*

*Proof.* When  $m = 1$ , the witness DFAs used in the proof of Theorem 1 in [23] can be used to show that the upper bound proposed in Theorem 8 can be reached.

Next, we consider the case when  $m \geq 2$ . We provide witness DFAs  $A$  and  $B$ , depicted in Figures 4 and 5, respectively, over the three letter alphabet  $\Sigma = \{a, b, c\}$ .

$A$  is defined as  $A = (Q_1, \Sigma, \delta_1, 0, \{0\})$  where  $Q_1 = \{0, 1, \dots, m-1\}$ , and the transitions are given as

- $\delta_1(i, a) = i + 1 \bmod m$ , for  $i \in Q_1$ ,
- $\delta_1(i, x) = i$ , for  $i \in Q_1$ , where  $x \in \{b, c\}$ .

$B$  is defined as  $B = (Q_2, \Sigma, \delta_2, 0, \{n-1\})$  where  $Q_2 = \{0, 1, \dots, n-1\}$ , where the transitions are given as

- $\delta_2(i, a) = i$ , for  $i \in Q_2$ ,
- $\delta_2(i, b) = i + 1 \bmod n$ , for  $i \in Q_2$ ,
- $\delta_2(0, c) = 0$ ,  $\delta_2(i, c) = i + 1 \bmod n$ , for  $i \in \{1, \dots, n-1\}$ .

Following the construction described in the proof of Theorem 8, we construct a DFA  $C = (Q, \Sigma, \delta, s, F)$  that accepts  $L(A)L(B)$  (also  $L(A)^*L(B)$ ). To prove that  $C$  is minimal, we show that (I) all the states in  $Q$  are reachable from  $s$ , and (II) any two different states in  $Q$  are not equivalent.

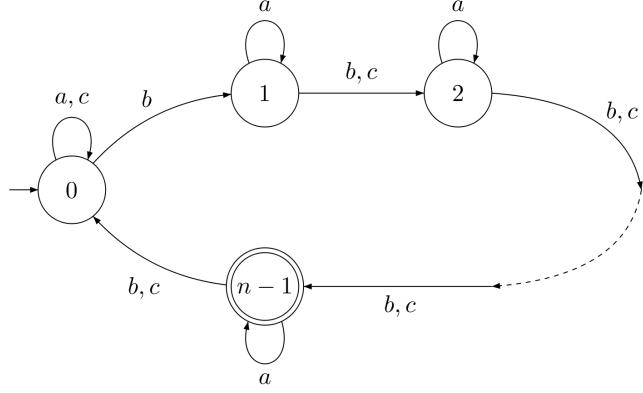


Figure 5: Witness DFA  $B$  for Theorem 9 when  $m \geq 2$

For (I), we show that all the states in  $Q$  are reachable by induction on the size of  $T$ .

The basis clearly holds, since, for any  $i \in Q_1$ , state  $\langle i, \{0\} \rangle$  is reachable from  $\langle 0, \{0\} \rangle$  by reading string  $a^i$ , and state  $\langle i, \{j\} \rangle$  can be reached from state  $\langle i, \{0\} \rangle$  on string  $b^j$ , for any  $i \in \{1, \dots, m-1\}$  and  $j \in Q_2$ .

In the induction steps, we assume that all the states  $\langle q, T \rangle$  such that  $|T| < k$  are reachable. Then, we consider the states  $\langle q, T \rangle$  where  $|T| = k$ . Let  $T = \{j_1, j_2, \dots, j_k\}$  such that  $0 \leq j_1 < j_2 < \dots < j_k \leq n-1$ . We consider the following three cases:

1.  $j_1 = 0$  and  $j_2 = 1$ . For any state  $i \in Q_1$ , state  $\langle i, T \rangle \in Q$  can be reached as

$$\langle i, \{0, 1, j_3, \dots, j_k\} \rangle = \delta(\langle 0, \{0, j_3-1, \dots, j_k-1\} \rangle, ba^i),$$

where  $\{0, j_3-1, \dots, j_k-1\}$  is of size  $k-1$ .

2.  $j_1 = 0$  and  $j_2 > 1$ . For any state  $i \in Q_1$ , state  $\langle i, \{0, j_2, \dots, j_k\} \rangle$  can be reached from state  $\langle i, \{0, 1, j_3-j_2+1, \dots, j_k-j_2+1\} \rangle$  by reading string  $c^{j_2-1}$ .

3.  $j_1 > 0$ . In such a case, the first component of state  $\langle q, T \rangle$  cannot be 0. Thus, for any state  $i \in \{1, \dots, m-1\}$ , state  $\langle i, \{j_1, j_2, \dots, j_k\} \rangle$  can be reached from state  $\langle i, \{0, j_2-j_1, \dots, j_k-j_1\} \rangle$  by reading string  $b^{j_1}$ .

Next, we show that any two distinct states  $\langle q, T \rangle$  and  $\langle q', T' \rangle$  in  $Q$  are not equivalent. We consider the following two cases:

1.  $q \neq q'$ . Without loss of generality, we assume  $q \neq 0$ . Then, string  $w = c^{n-1}a^{m-q}b^n$  can distinguish the two states, since  $\delta(\langle q, T \rangle, w) \in F$  and  $\delta(\langle q', T' \rangle, w) \notin F$ .

2.  $q = q'$  and  $T \neq T'$ . Without loss of generality, we assume that  $|T| \geq |T'|$ . Then, there exists a state  $j \in T - T'$ . It is clear that, when  $q \neq 0$ , string  $b^{n-1-j}$  can distinguish the two states, and when  $q = 0$ , string  $c^{n-1-j}$  can distinguish the two states since  $j$  cannot be 0.

Due to (I) and (II), DFA  $C$  needs at least  $m(2^n - 1) - 2^{n-1} + 1$  states and is minimal.  $\square$

In the rest of this section, we focus on the case where DFA  $A$  contains at least one final state that is not the initial state. Thus, this DFA is of size at least 2. We first obtain the following upper bound for the state complexity.

**Theorem 10.** *Let  $A = (Q_1, \Sigma, \delta_1, s_1, F_1)$  be a DFA such that  $|Q_1| = m > 1$  and  $|F_1 - \{s_1\}| = k_1 \geq 1$ , and  $B = (Q_2, \Sigma, \delta_2, s_2, F_2)$  be a DFA such that  $|Q_2| = n > 1$ . Then, there exists a DFA of at most  $(\frac{3}{4}2^m - 1)(2^n - 1) - (2^{m-1} - 2^{m-k_1-1})(2^{n-1} - 1)$  states that accepts  $L(A)^*L(B)$ .*

*Proof.* We denote  $F_1 - \{s_1\}$  by  $F_0$ . Then,  $|F_0| = k_1 \geq 1$ .

We construct a DFA  $C = \{Q, \Sigma, \delta, s, F\}$  for the language  $L_1^*L_2$ , where  $L_1$  and  $L_2$  are the languages accepted by DFAs  $A$  and  $B$ , respectively.

Let  $Q = \{\langle p, t \rangle \mid p \in P \text{ and } t \in T\} - \{\langle p', t' \rangle \mid p' \in P' \text{ and } t' \in T'\}$ , where

$$\begin{aligned} P &= \{R \mid R \subseteq (Q_1 - F_0) \text{ and } R \neq \emptyset\} \cup \{R \mid R \subseteq Q_1, s_1 \in R, \text{ and } R \cap F_0 \neq \emptyset\}, \\ T &= 2^{Q_2} - \{\emptyset\}, \\ P' &= \{R \mid R \subseteq Q_1, s_1 \in R, \text{ and } R \cap F_0 \neq \emptyset\}, \\ T' &= 2^{Q_2 - \{s_2\}} - \{\emptyset\}. \end{aligned}$$

The initial state  $s$  is  $s = \langle \{s_1\}, \{s_2\} \rangle$ .

The set of final states is defined to be  $F = \{\langle p, t \rangle \in Q \mid t \cap F_2 \neq \emptyset\}$ .

The transition relation  $\delta$  is defined as follows:

$$\delta(\langle p, t \rangle, a) = \begin{cases} \langle p', t' \rangle & \text{if } p' \cap F_1 = \emptyset, \\ \langle p', t' \cup \{s_2\} \rangle & \text{otherwise,} \end{cases}$$

where,  $a \in \Sigma$ ,  $p' = \delta_1(p, a)$ , and  $t' = \delta_2(t, a)$ .

Intuitively,  $C$  is equivalent to the NFA  $C'$  obtained by first constructing an NFA  $A'$  that accepts  $L_1^*$ , then catenating this new NFA with DFA  $B$  by  $\lambda$ -transitions. Note that, in the construction of  $A'$ , we need to add a new initial and final state  $s'_1$ . However, this new state does not appear in the first component of any of the states in  $Q$ . The reason is as follows. First, note that this new state does not have any incoming transitions. Thus, from the initial state  $s'_1$  of  $A'$ , after reading a nonempty word, we will never return to this state. As a result, states  $\langle p, t \rangle$  such that  $p \subseteq Q_1 \cup \{s'_1\}$ ,  $s'_1 \in p$ , and  $t \in 2^{Q_2}$  is never reached in DFA  $C$  except for the state  $\langle \{s'_1\}, \{s_2\} \rangle$ . Then, we note that, in the construction of  $A'$ , states  $s'_1$  and  $s_1$  should reach the same state on any letter in  $\Sigma$ . Thus, we can say that states  $\langle \{s'_1\}, \{s_2\} \rangle$  and  $\langle \{s_1\}, \{s_2\} \rangle$  are equivalent, because either

of them is final if  $s_2 \notin F_2$ , and they are both final states otherwise. Hence, we merge this two states and let  $\langle \{s_1\}, \{s_2\} \rangle$  be the initial state of  $C$ .

Also, we notice that states  $\langle p, \emptyset \rangle$  such that  $p \in P$  can never be reached in  $C$ , because  $B$  is complete.

Moreover,  $C$  does not contain those states whose first component contains a final state of  $A$  and whose second component does not contain the initial state of  $B$ .

Therefore, we can verify that DFA  $C$  indeed accepts  $L_1^* L_2$ , and it is clear that the size of  $Q$  is

$$\left(\frac{3}{4}2^m - 1\right)(2^n - 1) - (2^{m-1} - 2^{m-k_1-1})(2^{n-1} - 1).$$

□

Then, we show that this upper bound is reachable by some witness DFAs.

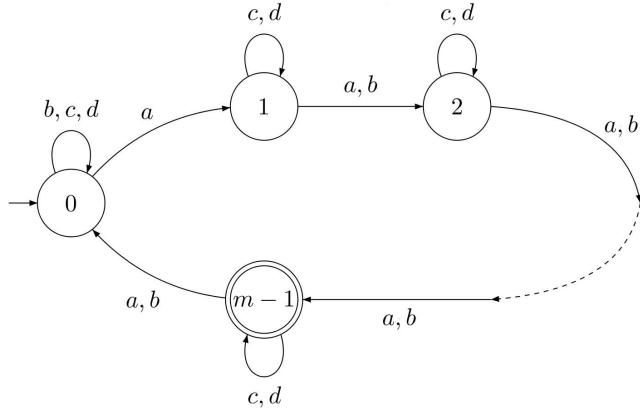


Figure 6: Witness DFA  $A$  for Theorem 11

**Theorem 11.** *For any integers  $m, n \geq 2$ , there exist a DFA  $A$  of  $m$  states and a DFA  $B$  of  $n$  states such that any DFA accepting  $L(A)^* L(B)$  needs at least  $5 \cdot 2^{m+n-3} - 2^{m-1} - 2^n + 1$  states.*

*Proof.* We define the following two automata over a four letter alphabet  $\Sigma = \{a, b, c, d\}$ .

Let  $A = (Q_1, \Sigma, \delta_1, 0, \{m-1\})$ , shown in Figure 6, where  $Q_1 = \{0, 1, \dots, m-1\}$ , and the transitions are defined as

- $\delta_1(i, a) = i + 1 \bmod m$ , for  $i \in Q_1$ ,
- $\delta_1(0, b) = 0$ ,  $\delta_1(i, b) = i + 1 \bmod m$ , for  $i \in \{1, \dots, m-1\}$ ,
- $\delta_1(i, x) = i$ , for  $i \in Q_1$ ,  $x \in \{c, d\}$ .

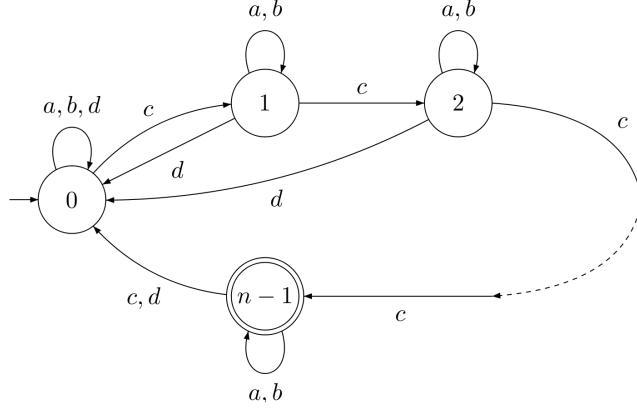


Figure 7: Witness DFA  $B$  for Theorem 11

Let  $B = (Q_2, \Sigma, \delta_2, 0, \{n-1\})$ , shown in Figure 7, where  $Q_2 = \{0, 1, \dots, n-1\}$ , and the transitions are defined as

- $\delta_2(i, x) = i$ , for  $i \in Q_2$ ,  $x \in \{a, b\}$ ,
- $\delta_2(i, c) = i + 1 \bmod n$ , for  $i \in Q_2$ ,
- $\delta_2(i, d) = 0$ , for  $i \in Q_2$ .

Let  $C = \{Q, \Sigma, \delta, \langle\{0\}, \{0\}\rangle, F\}$  be the DFA accepting the language  $L(A)^*L(B)$  which is constructed from  $A$  and  $B$  exactly as described in the proof of Theorem 10.

Now, we prove that the size of  $Q$  is minimal by showing that (I) any state in  $Q$  can be reached from the initial state, and (II) no two different states in  $Q$  are equivalent.

We first prove (I) by induction on the size of the second component  $t$  of the states in  $Q$ .

**Basis:** for any  $i \in Q_2$ , state  $\langle\{0\}, \{i\}\rangle$  can be reached from the initial state  $\langle\{0\}, \{0\}\rangle$  on string  $c^i$ . Then, by the proof of Theorem 5 in [23], it is clear that state  $\langle p, \{i\}\rangle$  of  $Q$ , where  $p \in P$  and  $i \in Q_2$ , is reachable from state  $\langle\{0\}, \{i\}\rangle$  on strings over letters  $a$  and  $b$ .

**Induction step:** assume that all the states  $\langle p, t \rangle$  in  $Q$  such that  $p \in P$  and  $|t| < k$  are reachable. Then, we consider the states  $\langle p, t \rangle$  in  $Q$  where  $p \in P$  and  $|t| = k$ . Let  $t = \{j_1, j_2, \dots, j_k\}$  such that  $0 \leq j_1 < j_2 < \dots < j_k \leq n-1$ .

Note that states such that  $p = \{0\}$  and  $j_1 = 0$  are reachable as follows:

$$\langle\{0\}, \{0, j_2, \dots, j_k\}\rangle = \delta(\langle\{0\}, \{0, j_3 - j_2, \dots, j_k - j_2\}\rangle, c^{j_2} a^{m-1} b).$$

Then, states such that  $p = \{0\}$  and  $j_1 > 0$  can be reached as follows:

$$\langle\{0\}, \{j_1, j_2, \dots, j_k\}\rangle = \delta(\langle\{0\}, \{0, j_2 - j_1, \dots, j_k - j_1\}\rangle, c^{j_1}).$$

Once again, by using the proof of Theorem 5 in [23], states  $\langle p, t \rangle$  in  $Q$ , where  $p \in P$  and  $|t| = k$ , can be reached from the state  $\langle \{0\}, t \rangle$  on strings over letters  $a$  and  $b$ .

Next, we show that any two states in  $Q$  are not equivalent. Let  $\langle p, t \rangle$  and  $\langle p', t' \rangle$  be two different states in  $Q$ . We consider the following two cases:

1.  $p \neq p'$ . Without loss of generality, we assume  $|p| \geq |p'|$ . Then, there exists a state  $i \in p - p'$ . It is clear that string  $a^{m-1-i}dc^n$  is accepted by  $C$  starting from state  $\langle p, t \rangle$ , but it is not accepted starting from state  $\langle p', t' \rangle$ .
2.  $p = p'$  and  $t \neq t'$ . We may assume that  $|t| \geq |t'|$  and let  $j \in t - t'$ . Then, state  $\langle p, t \rangle$  reaches a final state on string  $c^{n-1-j}$ , but state  $\langle p', t' \rangle$  does not on the same string. Note that, when  $m-1 \in p$ , we can say that  $j \neq 0$ .

Due to (I) and (II), DFA  $C$  has at least  $5 \cdot 2^{m+n-3} - 2^{m-1} - 2^n + 1$  reachable states, and any two of them are not equivalent.  $\square$

## 5 Conclusion

In this paper, we have studied the state complexities of two combined operations: reversal combined with catenation and star combined with catenation. We showed that, due to the structural properties of DFAs obtained from reversal and star, the state complexities of these two combined operations are not equal but close to the mathematical compositions of the state complexities of their individual participating operations.

## References

- [1] C. Campeanu, K. Culik, K. Salomaa, S. Yu: State complexity of basic operations on finite language, in: *Proceedings of the Fourth International Workshop on Implementing Automata VIII* 1-11, LNCS 2214, 1999, 60-70
- [2] C. Campeanu, K. Salomaa, S. Yu: Tight lower bound for the state complexity of shuffle of regular languages, *Journal of Automata, Languages and Combinatorics* 7 (3) (2002) 303-310
- [3] B. Cui, Y. Gao, L. Kari, S. Yu: State complexity of catenation combined with star and reversal, in: *Proceedings of DCFS 2010*, Saskatoon, SK, Canada, August 8-10, 2010
- [4] B. Cui, Y. Gao, L. Kari, S. Yu: State complexity of catenation combined with union and intersection, in: *Proceedings of CIAA 2010*, Winnipeg, MB, Canada, August 12-15, 2010
- [5] M. Daley, M. Domaratzki, K. Salomaa: State complexity of orthogonal catenation, in: *Proceedings of DCFS 2008*, Charlottetown, PE, Canada, July 16-18, 2008, 134-144

- [6] M. Domaratzki: State complexity and proportional removals, *Journal of Automata, Languages and Combinatorics* 7 (2002) 455-468
- [7] M. Domaratzki, A. Okhotin: State complexity of power, *Theoretical Computer Science* 410(24-25) (2009) 2377-2392
- [8] Z. Ésik, Y. Gao, G. Liu, S. Yu: Estimation of state complexity of combined operations, *Theoretical Computer Science* 410 (35) (2009) 3272-3280
- [9] Y. Gao, K. Salomaa, S. Yu: The state complexity of two combined operations: star of catenation and star of Reversal, *Fundamenta Informaticae* 83 (1-2) (2008) 75-89
- [10] Y. Gao and S. Yu: State complexity approximation, in: *Proceedings of Descriptive Complexity of Formal Systems* (2009) 163-174
- [11] Y. Gao and S. Yu: State complexity of union and intersection combined with star and reversal, *Computing Research Repository* (2010) arXiv:1006.3755v1
- [12] M. Holzer, M. Kutrib: State complexity of basic operations on non-deterministic finite automata, in: *Proceedings of International Conference on Implementation and Application of Automata 2002*, LNCS 2608, 2002, 148-157
- [13] J. E. Hopcroft, R. Motwani, J. D. Ullman: *Introduction to Automata Theory, Languages, and Computation (2nd Edition)*, Addison Wesley, 2001
- [14] J. Jirásek, G. Jirásková, A. Szabari: State complexity of concatenation and complementation of regular languages, *International Journal of Foundations of Computer Science* 16 (2005) 511-529
- [15] G. Jirásková: State complexity of some operations on binary regular languages, *Theoretical Computer Science* 330 (2005) 287-298
- [16] G. Jirásková, A. Okhotin: State complexity of cyclic shift, in: *Proceedings of DCFS 2005*, Como, Italy, June 30-July 2, 2005, 182-193
- [17] G. Jirásková, A. Okhotin: On the state complexity of star of union and star of intersection, *Turku Center for Computer Science TUCS Technical Report* No. 825, 2007
- [18] G. Liu, C. Martin-Vide, A. Salomaa, S. Yu: State complexity of basic language operations combined with reversal, *Information and Computation* 206 (2008) 1178-1186
- [19] G. Pighizzini, J. Shallit: Unary language operations, state complexity and Jacobsthal's function, *International Journal of Foundations of Computer Science* 13 (1) (2002) 145-159

- [20] A. Salomaa, K. Salomaa, S. Yu: State complexity of combined operations, *Theoretical Computer Science* 383 (2007) 140-152
- [21] A. Salomaa, D. Wood, S. Yu: On the state complexity of reversals of regular languages, *Theoretical Computer Science* 320 (2004) 293-313
- [22] S. Yu: State complexity of regular languages, *Journal of Automata, Languages and Combinatorics* 6 (2) (2001) 221-234
- [23] S. Yu, Q. Zhuang, K. Salomaa: The state complexity of some basic operations on regular languages, *Theoretical Computer Science* 125 (1994) 315-328
- [24] S. Yu: Regular languages, in: G. Rozenberg, A. Salomaa (Eds.), *Handbook of Formal Languages*, Vol. 1, Springer-Verlag, 1997, 41-110
- [25] S. Yu: State complexity of regular languages, *Journal of Automata, Languages and Combinatorics* 6(2) (2001) 221-234